Journal Acceptance Policies on ETDs

Statistics Department

California Polytechnic State University, San Luis Obispo

By Chelsea Kern

March 2013

Table of Contents

Overview…… 3

Project History…… 3

Background…… 4

Research Team…… 4

Frame Development…… 5

Pretesting…… 6

Sample Design…… 6

Data Collection……...7

Nonresponse Follow-up……8

Data Cleaning & Variable Creating……8

Nonresponse Bias……9

Science Analysis……10

Comparison Between Surveys……11

Conclusion……12

Reflection……13

Overview

I participated in survey research on science journals for my senior project. One of my first jobs was to clean and organize data files on our population of journals. I developed training notes and helped train students on conducting phone interviews. I created memos for the research team to carefully explain all statistical methods and analysis used in this project. Also, I cleaned the final data, imputed answers, and ran some basic analysis on our questions. Lastly, I have been participating in writing the manuscript that we hope to submit to The College & Research Libraries. We expect publication some time in 2014.

Project History

In the fall of 2011, Marisa Ramirez, a Digital Scholarship Services Librarian at Cal Poly, came to the statistical consulting service for advice on her research project, the goal of which had been to determine whether or not open access of Arts & Humanities electronic theses and dissertations did in fact diminish their future publishing opportunities. Marisa explained that when her research team sent their manuscript to a publisher, the publisher wanted to understand more about the how the study was conducted, and asked questions about nonresponse bias and the results of other statistical analyses that Marisa was unable to answer. Marisa wanted to know how to make her research project statistically sound, even though she had attempted a census of all social science & humanities journals.

Heather Smith, the statistical consultant at the time, helped Marisa answer the publisher's questions, and the publisher subsequently accepted the article after seeing the

supporting statistical data. Heather also shared with Marisa how she might have performed the entire study using statistical analysis tools, allowing for a more efficient study yet still delivering the most accurate results.

After the article was published, a question arose at a conference as to whether or not a similar study had been performed for science journals. Marisa returned to Heather for help with that new survey, and Heather asked if I wanted help with this study as my senior project.

Background

An increasing number of higher education institutions worldwide are requiring ETDs and are making them publicly available in open access repositories. Some faculty advisors and graduate students have expressed concern that open access to their electronic thesis or dissertation could diminish future publishing opportunities.

Research Team

A research group was formed that included: Marisa Ramirez, Digital Scholarship Services Librarian, Gail McMillan, Director of Digital Library and Archives, Joan Dalton, Associate University Librarian, Ann Halon, Digital Collections Coordinator, Professor Heather Smith, and myself. The research group wanted to know what the policies are for science journals acceptance of derivations of open access ETDs for publication, and whether or not the authors need be concerned about publishing opportunities.

Frame Development

Work began with the research team in May 2012. The team acquired an Excel file that identified 16,455 high impact science journals within 171 science disciplines using Thompson Reuter's Journal Performance Indicators (JPI). Because every science journal has multiple science disciplines that describe the various subjects of the journal, the JPI scores were used to determine journal importance within each discipline. A journal with a higher JPI score within the discipline is considered more relevant. We took the five publication titles with the highest JPI scores within each of the 171 science disciplines, for a total of 855 publication titles, and created a new Excel file with this information.

Journals appeared in our new Excel file one, two, three, or four times, depending on how many of the 171 science disciplines they appeared in within the top five JPI groupings. I removed 127 publication titles due to the multiple appearances, but made a note of how many times they appeared in the top 5 JPI rankings for use in the final selection process. I also removed five titles that were no longer published, leaving us with a data frame containing 723 journal titles.

I assigned the remaining 723 journals a subject category based on their title and science discipline. For example, "Journal of Engineering Education" fell into the engineering subject category, and "Emerging Infectious Disease" fell into the medical subject category. There were a total of 14 subjects including agriculture, biology, business/economics, chemistry, engineering, environmental science, math, medical, nutrition, ocean science, physics, psychology, transportation, and veterinary.

Pretesting

A pretest was conducted on eight purposefully selected journals. Each researcher chose to contact a few journal editors who they had some kind of connection to, or a subject that they were familiar with. We used the pretest interviews to verify that our main research question was worded clearly and our closed-ended options were adequate. We also learned that the best person to call about our main research question was the editor-in-chief or managing editor of each journal. Lastly, to no surprise, we were told that surveys get low priority, so we knew not to expect a very high response rate. A total of eight journals were pretested and taken out of our data frame in order to exclude them from possibly being in our sample.

Sample Design

After removing the eight journals that had been pretested, there were a total of 715 journals in our data frame, from which we took a sample of 300. The selection of the 300 journals was implemented using a stratification process, where a sample was taken from within each of the several strata. Our stratification variables included the 14 subject categories created, and the number of times a journal appeared in the top five JPI rankings. All 17 journals that appeared in the top five JPI rankings three or four times were included in the sample. Of the remaining 705 journals, we took a sample of 283 of them by taking roughly 40% of journals from each subject grouping. Below is a table that summarizes our stratified sampling plan.

Table 1. Summary of stratified sampling			
Stratum	**Number of journals**	**Sample size**	**Probability of Selection**
4 associated JPI disciplines	4	4	1
3 associated JPI disciplines	13	13	1
Agriculture	23	10	0.417
Biology	119	47	0.395
Business/Economics	17	7	0.412
Chemistry	70	28	0.400
Engineering	89	35	0.398
Environmental Science	73	29	0.403
Math	12	5	0.417
Medical	212	83	0.395
Nutrition	5	2	0.400
Ocean Science	25	10	0.400
Physics	52	21	0.420
Psychology	5	2	0.400
Transportation	5	2	0.400
Veterinary	4	2	0.500
Total	**715**	**300**	

Data Collection

Using our final sample, a contact file was developed in Excel that contained the name, email, and phone number of an editor from each journal. A survey script was also developed, the answers to which would identify the publishing policy for the sampled journals. The research team then sent out the link to our survey on Survey Monkey via email, with multiple follow up emails encouraging participation in our survey. After a few weeks, and 44 valid responses, the research team closed access to the survey.

Nonresponse Follow-up

Of the journals that did not respond, we took another sample of 100 journals in order to perform nonresponse follow-ups. Our goal was to see if the journals that had responded to our initial survey answered the ETD policy question systematically differently than those who did not respond. Four graduate students from University of Milwaukee-Wisconsin were paid to conduct phone call interviews on the sample of our non-respondents. I developed an interviewer-training sheet for the graduate students, which included background survey information, interview procedures, a pre-survey script, a voicemail script, a contact log, and the survey questionnaire. Professor Smith, Ann Hanlon, and I had a conference call on August 30th with the graduate students to explain the training materials and answer any questions they had. Through the phone interviews, we collected 28 responses over a two-week period.

Data Cleaning & Variable Creating

For our main research question:

"Manuscripts, which are revisions derived from openly accessible electronic theses or dissertations, are [considered]…" we created a new variable that we thought best described their policy based on what they answered and what types of comments they made about the question. For example, some people left the answer blank, but typed "always welcome" in the comment space. In that case, we imputed their answer for "ETD policy" from missing to "always welcome". Within this new variable, we also created new categories such as "not encountered" if the journal has never encountered this problem, "don't know" when the interviewee did not know the answer, and "not

applicable" when the journal considers themselves not applicable to this problem. Five of the 72 respondents had answers that fell into these new categories, and were removed from all further analyses regarding the main research question.

For journals that reported having an ETD policy, we grouped their ETD policies into a new variable with just two categories, "always welcome", and "restrictions". The "restrictions" category indicates that the journal's ETD policy was something other than "always welcome."

Nonresponse Bias

We used a Pearson Chi-Square test to determine whether or not there was evidence that the ETD policies were statistically different between the initial respondents and the respondents obtained during the follow-up interviews. Comparing policies that claimed ETDs were "always welcome" to those with "restrictions", we obtained a p-value of 0.101, which indicates (at an alpha = 0.5 level) that we do not have evidence that the non-respondents have different ETD policies than the respondents. Therefore we do not believe there is a non-response bias regarding the ETD policy question. There could possibly be a nonresponse bias, but because of our relatively small sample size, it would be very unlikely that we would have concluded there is a difference.

Table 2. Counts of ETD policy by r/nr

ETD Policy	r count	r %	nr count	nr %
Always Welcome	26	60%	11	46%
Restrictions	17	40%	13	54%
Total	**43**	**100%**	**24**	**100%**

Science Analysis

For those journals that reported having an ETD policy, we found that about 55% considered that manuscripts which are revisions derived from openly accessible ETDs are always welcome for submission, 21% are considered case-by-case, 9% are considered only if the contents and conclusions are substantially different from the ETD, 1% are considered only if the ETD has access limited to the institution where it was completed, and 13% are not considered under any circumstances. We are 95% confidence that between 43% and 67% of science journals consider open access ETDs always welcome for submission.

Table 3. ETD policies within science journals

ETD policy	Count	Percent
Always welcome	37	55%
Case by case	14	21%
Only if different	6	9%
Only if limited access	1	1%
Never	9	13%
Total	**67**	**100%**

Another question in our survey asked the survey respondent what their affiliation with their journal was. We performed a Pearson Chi-Square test to determine whether the ETD policies were statistically different between the affiliation types. We obtained a p-value of 0.11, which indicates (at an alpha = 0.5 level) that we do not have evidence to believe that the affiliation types have different ETD policies.

Table 4. Counts and percentages of ETD policy by affiliation type

Affiliation type	Always Welcome	Restrictions
Editor-in-chief	17 (71%)	19 (65%)
Managing editor	9 (26%)	8 (27%)
Assistant editor	6 (18%)	1 (3%)
Member of the editorial board	2 6%)	1 (3%)
Total	**34**	**29**

One question in our survey asked by what means their journal was published. We thought it would be interesting to see if ETD policies differed between publication types. Using the groupings "always welcome", and "restrictions", we performed a Pearson Chi-Square test to determine whether the ETD policies were statistically different between the publication types. We obtained a p-value of 0.161, which indicates (at an alpha = 0.5 level) that we do not have evidence to believe that the different publication types have different ETD policies.

Table 5. Counts of ETD policy by publication type

Publication Type	Always Welcome	Restrictions
Commercial Publishing Company	17 (47%)	19 (41%)
Academic Society	9 (52%)	8 (24%)
University Press	6 (86%)	1 (14%)
Total	**34**	**28**

We also looked at descriptive statistics, including the counts and percentages of the journals country of origin, ETD policy by subject category, counts and percentages of affiliation types, and counts and percentages of publication types.

Comparison Between Surveys

Next we wanted to compare the responses of our main research question between this survey of science journals and the previous survey of social science and humanities journals. I created a new variable for the social science and humanities ETD policy based on their answer and commentary, and also excluded the journals that answered "not encountered", "don't know", or "not applicable". For journals that reported having an ETD policy, we grouped their ETD policies into a new variable with three categories,

"always welcome", "some restrictions", and "never considered". The "some restrictions" category indicates that the journal's ETD policy was something other than "always welcome" or "never considered".

We found that 48% of arts & humanities journals answered "always welcome", 48% had some restrictions, and 4% said they were never considered under any circumstances. We used a Pearson Chi-Square test to determine whether or not the ETD policies are statistically different between the two surveys. We obtained a p-value of 0.025, which indicates (at an alpha = 0.05 level) that we have evidence that the ETD policies differ between science journals and social science and humanities journals. The "some restrictions" and "never considered" policies seems to be driving the differences between the two studies. Science journals tend to be a bit more restrictive than social science & humanities journals, likely due to the large number of medical journals that have restrictions.

Table 6. ETD policy by study type				
ETD policy	**Science counts**	**Science %**	**Social Science & Humanities counts**	**Social Science & Humanities %**
Always Welcome	37	55%	53	48%
Some Restrictions	21	31%	53	48%
Never Considered	9	13%	5	4%
Total	**67**	**100%**	**111**	**100%**

Conclusion

From this study, we can conclude that most science journals are open to considering manuscripts for publication that are revisions derived from openly accessible electronic theses and dissertations. In general, authors of ETDs need not be concerned

about publishing opportunities, but may benefit from researching the specific policies of the journals where they hope to be published.

Reflection

STAT421: Survey Sampling and Methodology provided a strong foundation for the process needed to conduct a successful analysis and answer the question about the effect of electronic dissertation publishing on future journal publishing opportunities. In addition to the importance of knowing proper analysis methods, professional presentation of data and documents is critical because it supports the credibility of information given and conclusions developed.

Working with a research team that never met in person made good communication a key aspect of the project. Finding a way to clearly explain statistical language and concepts to non-technical people is important to avoiding misunderstandings during the study process.

Developing a plan in advance for how to organize files, folders and data, so that it is easily accessible, helped to make the workflow significantly more efficient. Taking detailed notes and documenting the process as it happens made a big difference when I tried to recall and summarize later in the process. It also helped to avoid reworking data in order to record the steps taken in the study.

Having a well-planned and organized workflow, as well as an ability to clearly relate statistical concepts through written and oral communications, works hand-in-hand with understanding data when making it meaningful through statistical analysis.